On Finding a Subset of Non-Defective Items from a Large Population

Abhay Sharma and Chandra R. Murthy
Dept. of ECE, Indian Institute of Science, Bangalore 560 012, India
abhay.bits@gmail.com, cmurthy@iisc.ac.in

***Abstract*—In this paper, we derive mutual information based upper bounds on the number of nonadaptive group tests required to identify a given number of "non-defective" items from a large population containing a small number of "defective" items. In the asymptotic regime with the population size $N \to \infty$, to identify L non-defective items out of a population containing K defective items, our results show that $\frac{C_s K}{1-o(1)}(\Phi(\alpha_0,\beta_0)+o(1))$ measurements are sufficient when the tests are reliable. Here, C_s is a constant independent of N, K and L, and $\Phi(\alpha_0,\beta_0)$ is a bounded function of $\alpha_0 \triangleq \lim_{N\to\infty}\frac{L}{N-K}$ and $\beta_0 \triangleq \lim_{N\to\infty}\frac{K}{N-K}$. In contrast, the necessary number of tests using the conventional approach of first identifying the K defective items and picking the required number of nondefective items from the complement set grows with N as $O(K\log N)$. We also derive upper bounds on the number of tests under both dilution and additive noise models. Our results are obtained under a very general sparse signal model, by virtue of which, they are also applicable to other important sparse signal based applications such as compressive sensing.**

***Index Terms*—Sparse signal models, nonadaptive group testing, inactive subset recovery.**

I. Introduction

Sparse signal models are of great interest due to their applicability in a variety of areas such as group testing [2], [3], compressive sensing [4], signal de-noising [5], subset selection [6], etc. Generally speaking, in a sparse signal model, out of a given number N of input variables, only a small subset of size K contributes to the observed output. For example, in a non-adaptive group testing setup, the output depends only on whether the items from the defective set participate or do not participate in the group test. Similarly, in a compressive sensing setup, the output signal is a set of random projections of the signal corresponding to the non-zero entries (support set) of the input vector. This *salient* subset of inputs is referred to by different names, e.g., defective items, sick individuals, support set, etc. In the sequel, we will refer to it as *the active set*, and its complement as *the inactive set*. In this paper, we address the issue of the *inactive subset recovery*. That is, we focus on the task of finding an L $(\leq N-K)$ sized subset of *the inactive set* (of size $N-K$), given the observations from a sparse signal model with N inputs, out of which K are active.

The problem of finding a subset of items belonging to the inactive set is of interest in many applications. An example is the spectrum hole search problem in the cognitive radio (CR) networks [7]. It is well known that the primary user occupancy (active set) is sparse in the frequency domain over a wide band of interest [8], [9]. To setup a CR network, the secondary users need to find an appropriately wide unoccupied (inactive) frequency band. Thus, the main interest here is the identification of *only a sub-band* out of the total available unoccupied band, i.e., it is an inactive subset recovery problem. Furthermore, the required bandwidth of the spectrum hole will typically be a small fraction of the entire bandwidth that is free at any point in time [10]. Another example is a product manufacturing plant, where a small shipment of non-defective (inactive) items has to be delivered on high priority. Once again, it is of interest to identify a subset of the non-defective items using as few tests as possible.

Related work: In the group testing literature, the problem of bounding the number of tests required to identify the defective items in a large population has been studied, both in the noiseless and noisy settings, for tractable decoding algorithms as well as under general information theoretic models [11]–[21]. A combinatorial approach has been adopted in [11], [12], [22], where explicit constructions for the test matrices are used, e.g., using superimposed codes, to design matrices with properties that ensure guaranteed detection of a small number of defective items. Two such properties were considered: disjunctness and separability [3].[1] A probabilistic approach was adopted in [13]–[15], [23], where random test matrix designs were considered, and upper and lower bounds on the number of tests required to satisfy the properties of disjunctness or separability with high probability were derived. Another study [20] uses random test designs, and develops computationally efficient algorithms for identifying defective items from the noisy test outcomes by exploiting the connection with compressive sensing. An approach based on information density is used in [21] to analyze the phase transition behavior of Bernoulli test matrix designs and propose measurement-optimal recovery algorithms. A general sparse signal model for studying group testing problems, that turns out to be very useful in dealing with noisy settings, was proposed and used in [16]–[19]. In this framework, the group testing problem was formulated as a detection problem and a one-to-one correspondence was established with a communication

This work was presented in part in [1]. It was financially supported by a research grant and fellowship from the Ministry of Electronics and Information Technology, Govt. of India.

[1] A test matrix, with tests indexing the rows and items indexing the columns, is said to be k-disjunct if the boolean sum of every k columns does not equal any other column in the matrix. Also, a test matrix is said to be k-separable if the boolean sum of every set of k columns is unique.

channel model. Using information theoretic arguments, mutual information based expressions (that are easily computable for a variety of noise models) for upper and lower bounds on the number of tests were obtained [19].

The problem of non-defective subset identification can be related to the problem of group testing using list decoding [24]–[26], where the decoder outputs a superset of the true defective set, i.e., a list of items $\mathcal{L}$ (with $|\mathcal{L}| > K$) such that $\mathcal{L}$ contains the defective set. It finds applications in scenarios where it is permissible for some non-defective items to be included in the decoded set, as long as it contains most of the defective items. For example, the output of the list decoder could be used as a first step in a two-stage decoding procedure in group testing, fault detection applications, etc. In this setup, in contrast to our problem, the typical regimes of interest are those where the list size, although larger than K, is still comparable to K and much smaller than N. In [25], list decoding has been studied as an intermediate step in conventional group testing decoding. A combinatorial approach employing list-disjunct matrices was used to derive bounds on number of tests. A very recent work [26] studies list-decoding with partial recovery under the scaling regimes $|\mathcal{L}| = o(N)$. The authors show that while list decoding may offer significant benefit when $|\mathcal{L}| = O(N)$ (which is shown in [27] in the context of non-defective subset recovery), the gains are limited in the $|\mathcal{L}| = o(N)$ regime. Another recent work [28] studies the problem of finding zeros in a sparse vector in the framework of compressive sensing. The authors propose computationally efficient recovery algorithms and study their performance through simulations.

In this paper, we build on [1] and focus on deriving **information theoretic upper bounds (i.e., sufficient conditions)** on the number of measurements needed for identifying a given number of inactive items in a large population with arbitrarily small probability of error. We consider the general sparse signal model employed in [16], [19] in the context of the support recovery (i.e., defective set recovery) problem. The model consists of N input covariates, out of which, an unknown subset S of size K are “active”. Only the active variables, i.e., the variables from the set S, are relevant to the output. Mathematically, this is modeled by assuming that, given the active set S, the output Y is independent of remaining input variables. Further, the probability distribution of the output conditioned on a given active set, is assumed to be known for all possible active sets. Given multiple observations from the this model, we propose and analyze decoding schemes to identify *a set of L inactive variables*. We compare two alternative decoding schemes: (a) Identify the active set and then choose L inactive covariates randomly from the complement set, and, (b) Decode the inactive subset directly from the observations. Our main contributions are as follows:

1) We analyze the average probability of error for both the decoding schemes. We use the analysis to obtain mutual information based upper bounds on the number of observations required to identify a set of L inactive variables with the probability of error decreasing exponentially with the number of observations.
2) We specialize the above bounds to various noisy non-adaptive group testing scenarios, and characterize the number of tests required to identify L non-defective items, in terms of L, N and K.

Our results show that, compared to the conventional approach of identifying the inactive subset by first identifying the active set, directly searching for an L-sized inactive subset offers a reduction in the number of observations (tests/measurements), which is especially significant when L is small compared to $N-K$. When the tests are reliable, in the asymptotic regime as $N \to \infty$, if $\frac{L}{N-K} \to \alpha_0$ and $\frac{K}{N-K} \to \beta_0$, $\frac{C_s K}{1-o(1)}(\Phi(\alpha_0, \beta_0) + o(1))$ measurements are sufficient, where C_s is a constant independent of N, K and L, and $\Phi(\alpha_0, \beta_0)$ is a bounded function of α_0 and β_0. We show that this improves on the number of observations required by the conventional approach, in the sequel.

The rest of the paper is organized as follows. Section II describes the signal model and problem setup. We present our upper bounds on the number of observations in Sections III. An application of the bounds to group testing is described in Section IV. The proofs for the main results are provided in Section V, and concluding remarks are offered in Section VI.

Notation: For any positive integer a, $[a] \triangleq \{1, 2, \ldots, a\}$. For any set A, A^c denotes complement operation and $|A|$ denotes the cardinality of the set. For any two sets A and B, $A \backslash B = A \cap B^c$, i.e., elements of A that are not in B. $\{\emptyset\}$ denotes the null set. Scalar random variables (RVs) are represented by capital non-bold alphabets, e.g., $\{Z_1, Z_3, Z_5, Z_8\}$ represent a set of 4 scalar RVs. If the index set is known, we also use the index set as a subscript, e.g., Z_S, where $S = \{1, 3, 5, 8\}$. Bold-face letters represent random matrices (or a set of vector random variables). We use an index set to specify a subset of columns from the given random matrix. For example, let $\mathbf{Z}$ denote a random matrix with n columns. For any $S \subset [n]$, $\mathbf{Z}_S$ denotes a set of $|S|$ columns of $\mathbf{Z}$ specified by the index set S. Individual vector RVs are also denoted using an underline, e.g., $\underline{\mathbf{z}}$ represents a single random vector. For any discrete random variable Z, $\{Z\}$ represents the set of all realizations of Z. Similarly, for a random matrix $\mathbf{Z}$, whose entries are discrete random variables, $\{\mathbf{Z}\}$ represents the set of all realizations of $\mathbf{Z}$. For any two jointly distributed random variables $\underline{\mathbf{z}}_1$ and $\underline{\mathbf{z}}_2$, with a slight abuse of notation, let $P(\underline{\mathbf{z}}_1|\underline{\mathbf{z}}_2)$ denote the conditional probability distribution of $\underline{\mathbf{z}}_1$ given “a realization $\underline{\mathbf{z}}_2$” of the random variable $\underline{\mathbf{z}}_2$. Similarly, $P(\underline{\mathbf{z}}_1|\mathbf{Z})$ denotes the conditional probability distribution of $\underline{\mathbf{z}}_1$, given a realization $\mathbf{Z}$ of the random matrix $\mathbf{Z}$. $\mathcal{B}(q), q \in [0\ 1]$ denotes the Bernoulli distribution with parameter q. $\mathbb{I}_{\mathcal{A}}$ denotes the indicator function, which returns 1 if the event $\mathcal{A}$ is true, and returns 0 otherwise. Note that, $x(n) = O(y(n))$ implies that $\exists\ B > 0$ and $n_0 > 0$, such that $|x(n)| \leq B|y(n)|$ for all $n > n_0$. Similarly, $x(n) = \Omega(y(n))$ implies that $\exists\ B > 0$ and $n_0 > 0$, such that $|x(n)| \geq B|y(n)|$ for all $n > n_0$. Also, $x(n) = o(y(n))$ implies that for every $\epsilon > 0$, there exists an $n_0 > 0$ such that $|x(n)| \leq \epsilon|y(n)|$ for all $n > n_0$. In this work, unless otherwise specified, all logarithms are to the base e. For any $p \in [0, 1]$, $H_b(p)$ denotes the binary entropy in nats, i.e., $H_b(p) \triangleq -p\log(p) - (1-p)\log(1-p)$.

II. Problem Setup

In this section, we describe the signal model and problem setup. Let $X_{[N]} = \left[X_1, X_2, \ldots, X_N\right]$ denote a set of N independent and identically distributed (i.i.d.) input random variables (or *items*). Let each X_j belong to a finite alphabet denoted by $\mathcal{X}$ and be distributed as $\Pr\{X_j = x\} = Q(x), x \in \mathcal{X},\ j = 1, 2, \ldots, N$. For a group of input variables, e.g., $X_{[N]}$, $Q(X_{[N]}) = \prod_{j\in[N]} Q(X_j)$ denotes the known joint distribution for all the input variables. We consider a sparse signal model where only a subset of the input variables are *active* (or *defective*), in the sense that only a subset of the input variables contribute to the output. Let $S \subset [N]$ denote the set of input variables that are active, with $|S| = K$. We assume that K, i.e., the size of the active set, is known. Let $S^c \triangleq [N] \backslash S$ denote the set of variables that are *inactive* (or *non-defective*). Let the output belong to a finite alphabet denoted by $\mathcal{Y}$. We assume that Y is generated according to a known conditional distribution $P(Y|X_{[N]})$. Then, in our observation model, we assume that given the active set, S, the output signal, Y, is independent of the other input variables. That is, $\forall\ Y \in \mathcal{Y}$,

$$P(Y|X_{[N]}) = P(Y|X_S). \tag{1}$$

We observe the outputs corresponding to M independent realizations of the input variables, and denote the inputs and the corresponding observations by $\{\mathbf{X}, \underline{\mathbf{y}}\}$. Here, $\mathbf{X}$ is an $M \times N$ matrix, with its i^{th} row representing the i^{th} realization of the input variables, and $\underline{\mathbf{y}}$ is an $M \times 1$ vector, with its i^{th} component representing the i^{th} observed output. Note that, the independence assumption across the input variables and across different observations implies that each entry in $\mathbf{X}$ is i.i.d. Let $L \leq N - K$. We consider the problem of finding *a set* of L inactive variables given the observation set, $\{\mathbf{X}, \underline{\mathbf{y}}\}$. That is, we wish to find an index set $S_H \subset S^c$ such that $|S_H| = L$. In particular, our goal is to derive information theoretic bounds on the number of observations (measurements/group tests) required to find a set of L inactive variables with the probability of error exponentially decreasing with the number of observations. Here, an error event occurs if the chosen inactive set contains one or more active variables. Now, one way to find L inactive variables is to find all the active variables and then choose any L variables from the complement set. Thus, existing bounds on M for finding the active set are an upper bound on the number of observations required for solving our problem. However, intuitively speaking, fewer observations should suffice to find L inactive variables, since we do not need to find the full active set. This is confirmed by our results presented in the next section.

The above signal model can be equivalently described using Shannon's random codebook based channel coding framework; see Figure 1. The active set S corresponds to one of the $\binom{N}{K}$ possible active sets with K variables, and constitutes the input message. Let $\mathbf{X} \in \mathcal{X}^{M\times N}$ be a random codebook consisting of N codewords of length M; each associated with one of the N input variables. Let $\underline{\mathbf{x}}_i$ denote the codeword associated with i^{th} input variable. The encoder encodes the message as a length-M message $\mathbf{X}_S \in \mathcal{X}^{M\times K}$, that comprises of K codewords, each of length M, chosen according to the index set S from $\mathbf{X}$. That is, $\mathbf{X}_S = [\underline{\mathbf{x}}_{i_1}\ \underline{\mathbf{x}}_{i_2} \ldots \underline{\mathbf{x}}_{i_K}]$, for each $i_l \in S$. Let $\mathbf{X}_S^{(i)}$ denote the i^{th} row of the matrix $\mathbf{X}_S$ and let $\underline{\mathbf{y}}(i)$ denote its i^{th} component. The encoded message is transmitted through a discrete memoryless channel [29], [30], denoted by $(\mathcal{X}^M, P(\underline{\mathbf{y}}|\mathbf{X}_S), \mathcal{Y}^M)$, where $P(\underline{\mathbf{y}}|\mathbf{X}_S) = \prod_{i=1}^{M} P(\underline{\mathbf{y}}(i)|\mathbf{X}_S^{(i)})$ and the distribution function $P(\underline{\mathbf{y}}(i)|\mathbf{X}_S^{(i)})$ is known for each active set S. Given the codebook $\mathbf{X}$ and the output message $\underline{\mathbf{y}}$, our goal is to find *a set* of L variables *not* belonging to the active set S. Also, the above signal model, proposed and used earlier in [16], [19], is a generalization of the signal models employed in some of the popular non-adaptive measurement models such as compressed sensing[2] and non-adaptive group testing. Thus, the general mutual information based bounds on number of observations to find a set of inactive items obtained using the above model are applicable in a variety of practical scenarios.

We now discuss the above signal model in context of a specific non-adaptive measurement system, namely the random pooling based, noisy non-adaptive group testing framework [3], [19]. In a group testing framework [3], [16], [19], we have a population of N items, out of which K are defective. Let $\mathcal{G} \subset [N]$ denote the defective set, such that $|\mathcal{G}| = K$. The group tests are defined by a boolean matrix, $\mathbf{X} \in \{0,1\}^{M\times N}$, that assigns different items to the M group tests (pools). In the i^{th} test, the items corresponding to the columns with 1 in the i^{th} row of $\mathbf{X}$ are tested. As in [19], we consider an i.i.d. random Bernoulli measurement matrix, where each $X_{ij} \sim \mathcal{B}(p)$ for some $0 < p < 1$. Here, p is a design parameter that controls the average group size. If the tests are completely reliable, then the output of the M tests is given by the boolean OR of the columns of $\mathbf{X}$ corresponding to the *defective set* $\mathcal{G}$. We consider the following two noise models [15], [19]: (a) An *additive* noise model, where there is a probability, $q \in (0, 1]$, that the outcome of a group test containing only non-defective items comes out positive; (b) A *dilution* model, where there is a probability, $u \in (0, 1]$, that a given item does not participate in a given group test. Let $\underline{d}_i \in \{0,1\}^M$. Let $\underline{d}_i(j) \sim \mathcal{B}(1-u)$ be chosen independently for all $j = 1, 2, \ldots M$ and for all $i = 1, 2, \ldots N$. Let $\mathbf{D}_i \triangleq \text{diag}(\underline{d}_i)$. Let "$\bigvee$" denote the boolean OR operation. The output vector $\underline{y} \in \{0,1\}^M$ can be represented as

$$\underline{y} = \bigvee_{i\in\mathcal{G}} \mathbf{D}_i \underline{x}_i \bigvee \underline{w}, \tag{2}$$

where $\underline{x}_i \in \{0,1\}^M$ is the i^{th} column of $\mathbf{X}$, $\underline{w} \in \{0,1\}^M$ is the additive noise with the i^{th} component $\underline{w}(i) \sim \mathcal{B}(q)$. For the noiseless case, $u = 0, q = 0$. In an additive model, $u = 0, q > 0$. In a dilution model, $u > 0, q = 0$. This "logical-OR" signal model captures many practical non-adaptive group testing measurement systems, see, e.g., [3], [15], [33], [34].

We now relate this model with the general sparse signal model described above. Note that, $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1\}$. Each item in the group testing framework corresponds to one of the N input covariates. The i^{th} *row* of the test matrix, which

[2] Although we focus on models with finite alphabets in this work, our results easily extend to models with continuous alphabets [31], [32].

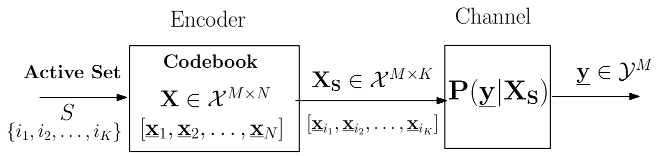
Encoder
Channel
Active Set
S
$\{i_1, i_2, \ldots, i_K\}$
Codebook
$\mathbf{X} \in \mathcal{X}^{M\times N}$
$[\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_N]$
$\mathbf{X}_\mathbf{S} \in \mathcal{X}^{M\times K}$
$[\underline{\mathbf{x}}_{i_1}, \underline{\mathbf{x}}_{i_2}, \ldots, \underline{\mathbf{x}}_{i_K}]$
$\mathbf{P}(\underline{\mathbf{y}}|\mathbf{X}_\mathbf{S})$
$\underline{\mathbf{y}} \in \mathcal{Y}^M$

Fig. 1. Sparse signal model: An equivalent random codebook based channel coding model.

specifies the i^{th} random pool, corresponds to the i^{th} realization of the input covariates. From (2), given the defective set $\mathcal{G}$, the i^{th} test outcome $\underline{\mathbf{y}}(i)$ is independent of values of input variables from the set $[N]\backslash\mathcal{G}$. That is, with regards to test outcome, it is *irrelevant* whether the items from the set $[N]\backslash\mathcal{G}$ are included in the test or not. Thus, $\mathcal{G}$ corresponds to the active set S. Further, with regards to the channel coding setup, the test matrix $\mathbf{X}$ corresponds to the random codebook, and each column specifies the M length random code with the associated item. The channel model, i.e., the probability distribution functions $P(\underline{\mathbf{y}}|\mathbf{X}_\mathcal{G})$ for any $\mathcal{G}$, is fully determined from (2) and the statistical models for the dilution and additive noise. Thus, it is easy to see that the group testing framework is a special case of the general sparse model that we have considered, and, the number of group tests correspond directly to the number of observations in the context of sparse models.

We now define two quantities that are very useful in the development to follow. Let S be a given active set. For any $1 \le j \le K$, let $S^{(j)}$ and $S^{(K-j)}$ represent a partition of S such that $S^{(j)} \cup S^{(K-j)} = S$, $S^{(j)} \cap S^{(K-j)} = \{\emptyset\}$ and $|S^{(j)}| = j$. Define

$$E_0(\rho, j, n) = -\log \sum_{Y\in\mathcal{Y}} \sum_{X_{S^{(K-j)}}\in\mathcal{X}^{K-j}} \left\{ \sum_{X_{S^{(j)}}\in\mathcal{X}^j} Q(X_{S^{(j)}}) \left(P(Y, X_{S^{(K-j)}}|X_{S^{(j)}})\right)^{\frac{1}{1+\rho n}} \right\}^{1+\rho n} \tag{3}$$

for any positive integer n and any $\rho \in [0, 1]$. Also, let $I^{(j)} \triangleq I(Y, X_{S^{(K-j)}}; X_{S^{(j)}}) = I(Y; X_{S^{(j)}}|X_{S^{(K-j)}})$ be the mutual information between $\{Y, X_{S^{(K-j)}}\}$ and $X_{S^{(j)}}$ [29], [30]. Mathematically,

$$I^{(j)} = \sum_{Y\in\mathcal{Y}} \sum_{X_{S^{(K-j)}}\in\mathcal{X}^{K-j}} \sum_{X_{S^{(j)}}\in\mathcal{X}^j} P(Y, X_{S^{(K-j)}}|X_{S^{(j)}}) Q(X_{S^{(j)}}) \log \frac{P(Y, X_{S^{(K-j)}}|X_{S^{(j)}})}{P(Y, X_{S^{(K-j)}})}. \tag{4}$$

Using the independence assumptions in the signal model, by the symmetry of the codebook construction, for a given j, $E_0(\rho, j, n)$ and $I^{(j)}$ are independent of the specific choice of S, and of the specific partitions of S. It is easy to verify that $\frac{dE_0(\rho, j, n)}{d\rho}|_{\rho=0} = nI^{(j)}$. Furthermore, it can be shown that $E_0(\rho, j, n)$ is a concave function of ρ [29] (also see Figure 4).

III. Sufficient Number of Observations

We first present results on the sufficient number of observations to find a set of L inactive variables. The general methodology used to find the upper bounds is as follows: (a) Given a set of inputs and observations, $\{\mathbf{X}, \underline{\mathbf{y}}\}$, we first propose a decoding algorithm to find an L-sized inactive set, S_H; (b) For the given decoding scheme, we find (or upper bound) the average probability of error, where the error probability is averaged over the random set $\{\mathbf{X}, \underline{\mathbf{y}}\}$ as well as over all possible choices for the active set. An error occurs when the decoded set of L inactive variables contains one or more active variables. That is, with S as the active set and S_H as the *decoded* inactive set, an error occurs if $S \cap S_H \neq \{\emptyset\}$; (c) We find the relationships between M, N, L and K that will drive the average probability of error to zero. Section III-A describes the straightforward decoding scheme where we find the inactive variables by first isolating the active set followed by choosing the inactive set randomly from the complement set. This is followed by the analysis of a new decoding scheme we propose in Section III-B, where we directly search for an inactive subset of the required cardinality.

A. Decoding scheme 1: Look into the Complement Set

One way to find a set of inactive (or non-defective) variables is to first decode the active (defective) set and then pick a set of L variables uniformly at random from the complement set. Here, we employ maximum likelihood based optimal decoding [19] to find the active set. Intuitively, even if we choose a wrong active set, there is still a nonzero probability of picking a correct inactive set, since there remain only a few active variables in the complement set. We refer to this decoding scheme as the "indirect" decoding scheme. The probability of error in identifying the active set was analyzed in [19]. The error probability when the same decoding scheme is employed to identify a inactive subset is similar, with an extra term to account for the probability of picking an incorrect set of L variables from the complement set. For this decoding scheme, we present the following result, without proof, as a corollary to (Lemma III.I, [19]).

Corollary 1. *Let N, M, L and K be as defined above. For any $\rho \in [0, 1]$, with the above decoding scheme, the average probability of error, P_e, in finding L inactive variables is upper bounded as*

$$P_e \le \max_{1\le j\le K} \exp\left\{-\left(ME_0(\rho, j, 1) - \rho\log\left[\binom{N-K}{j} C_0(j)\right] - \log\left[K\binom{K}{j}\right]\right)\right\}, \tag{5}$$

where $C_0(j) \triangleq \frac{\sum_{i=1}^{j}\binom{N-K-j}{L-i}\binom{j}{i}}{\binom{N-K}{L}}$ denotes the probability of error in choosing L inactive variables uniformly at random from $N-K$ variables containing j active variables.

From above, by lower bounding $E_0(\rho, j, 1)$ for any specific signal model, we can obtain a bound that gives us the sufficient number of observations to find a set of L inactive variables. We obtain the corresponding bound in the context of non-adaptive group testing in Section IV (see Corollary 2). Since $C_0 \le 1$, this bound is tighter than the bound obtained by using the same number of observations as is required to find the active set [19].

B. Decoding Scheme 2: Find the Inactive Subset Directly

For simplicity of exposition, we describe this decoding scheme in two stages: First, we present the result for the $K = 1$ case, i.e., when there is only one active variable. This case brings out the fundamental difference between finding active and inactive variables. We then generalize our decoding scheme to $K > 1$.

1) The $K = 1$ Case

We start by proposing the following decoding scheme:

- Given $\{\mathbf{X}, \underline{\mathbf{y}}\}$, compute $P(\underline{\mathbf{y}}|\underline{\mathbf{x}}_i)$ for all $i \in [N]$ and sort them in descending order. Since $K = 1$, we know $P(Y|X_i)$ for all $i \in [N]$, and hence $P(\underline{\mathbf{y}}|\underline{\mathbf{x}}_i)$ can be computed using the independence assumption across different observations.
- Pick the last L indices in the sorted array as the set of L inactive variables.

Note that, in contrast to finding active set, the problem of finding L inactive variables does not have unique solution (except for $L = N - K$). The proposed decoding scheme provides a way to pick a solution, and the probability of error analysis takes into account the fact that an error event happens only when the inactive set chosen by the decoding algorithm contains an active variable.

Theorem 1. *Let N, M, L and K be as defined above. Let $K = 1$. Let E_0 and $I^{(j)}$ be as defined in (3) and (4). Let $\rho \in [0\ 1]$. With the above decoding scheme, the average probability of error, P_e, in finding L inactive variables is upper bounded as*

$$P_e \leq \exp\left[-\left(ME_0(\rho, 1, N-L) - \rho \log \binom{N-1}{L-1}\right)\right]. \tag{6}$$

Further, for any $\epsilon_0 > 0$, if

$$M > (1+\epsilon_0)\frac{\log\binom{N-1}{L-1}}{(N-L)I^{(1)}}, \tag{7}$$

then there exists $\epsilon_1 > 0$, independent of N and L, such that $P_e \leq \exp\left(-\epsilon_1 \log\binom{N-1}{L-1}\right)$.

Proof: *See Sec. V-A.*

We make the following observations:

(a) Figure 2 compares the above bound on the number of observations with the bounds for the decoding scheme presented in Section III-A[3] and in Theorem III.I [19], for $K = 1$.

(b) Consider the case $L = N - 1$, i.e., we want to find all the inactive variables. This task is equivalent to finding the active variable. The above decoding scheme for finding $N - 1$ inactive variables is equivalent[4] to the maximum likelihood criterion based decoding scheme used in Theorem III.I in [19] for finding 1 active variable. This is also reflected in the above result, as the number of observations sufficient for finding $N - 1$ inactive variables matches exactly with the number of observations sufficient for finding 1 active variable (see Figure 2).

Intuitively, out of the $\binom{N}{L}$ possible sets of size L, $\binom{N-1}{N-L-1}$ contain only inactive variables. Thus, $\log\left(\binom{N}{L}/\binom{N-1}{N-L-1}\right)$ number of bits can describe all the sets corresponding to each "right choice," i.e., corresponding to each L-sized set containing only inactive variables. Since $I^{(1)}$ denotes the amount of information obtained per observation, $\frac{\log\frac{N}{N-L}}{I^{(1)}} \approx \frac{\log\binom{N-1}{L-1}}{(N-L)I^{(1)}}$ equals the number of observations required for finding an inactive set. Hence, the result in Theorem 1 is intuitively satisfying.

2) $K > 1$ Case

For $K > 1$, by arranging $P(\underline{\mathbf{y}}|\mathbf{X}_{S_i})$ in decreasing order for all $S_i \subset [N]$ such that $|S_i| = K$, it is possible for the sets S_i towards the end of the sorted list to have overlapping entries. Thus, in this case the decoding algorithm proceeds by picking up just the sufficient number of K-sized sets from the end that provides us with a set of L inactive variables. We propose the following decoding scheme:

Decoding Scheme:

1) Given $\{\mathbf{X}, \underline{\mathbf{y}}\}$, compute $P(\underline{\mathbf{y}}|\mathbf{X}_{S_i})$ for all $S_i \subset [N]$ such that $|S_i| = K$, and sort these in descending order. Let the ordering be denoted by $\{S_{i_1}, S_{i_2}, \ldots, S_{i_{\binom{N}{K}}}\}$.
2) Choose n_0 sets from the end such that
$$|\bigcup_{l=1}^{n_0} S_{i_{\binom{N}{K}-l+1}}| \geq L \quad \text{and} \quad |\bigcup_{l=1}^{n_0-1} S_{i_{\binom{N}{K}-l+1}}| < L. \tag{8}$$
3) Let $\Omega_{\text{last}} \triangleq \{i_{\binom{N}{K}}, i_{\binom{N}{K}-1}, \ldots, i_{\binom{N}{K}-n_0+1}\}$ denote this set of last n_0 indices. Declare $S_H \triangleq \bigcup_{j \in \Omega_{\text{last}}} S_j$ as the decoded set of inactive variables.

That is, choose the minimum number of K-sized sets with least likelihoods such that we get L distinct variables and declare these as the decoded set of inactive variables. We refer to this decoding scheme as the "direct" decoding scheme. We note that S_H might contain more than L items. In particular, $L \leq |S_H| \leq L + K - 1$. Further, for all values of L such that $L < (N-K) - (K-1)$, the complement set of S_H, i.e., $[N]\backslash S_H$, will contain at least $L_0 \triangleq (N - L - 2K + 1)$ variables from the inactive set $[N]\backslash S_1$. This will be useful in deriving an upper bound on the decoding error probability for this algorithm. We summarize the probability of error analysis of the above algorithm in the following theorem.

Theorem 2. *Let N, M, L and K be as defined above. Let $L_0 \triangleq (N - L - 2K + 1)$. For any $\rho \in [0\ 1]$ and any $1 \leq L < (N-K) - (K-1)$, with the above decoding scheme, the average probability of error, P_e, in finding L inactive variables is upper bounded as*

$$P_e \leq \exp\left[-\left\{ME_0(\rho, 1, L_0) - \rho \log \binom{N-K}{L_0} - \log\left[K\binom{N-1-L_0}{K-1}\right]\right\}\right]. \tag{9}$$

Proof: See Sec. V-B.

[3] We refer the reader to the remark at the end of the proof for Theorem 1 (Section V-A) for a bound on the sufficient number of observations, resulting from Corollary 1, corresponding to $K = 1$ case.

[4] The decoding schemes are equivalent in the sense that an error in finding K active variables implies an error in finding $N - K$ inactive variables, and vice-versa.

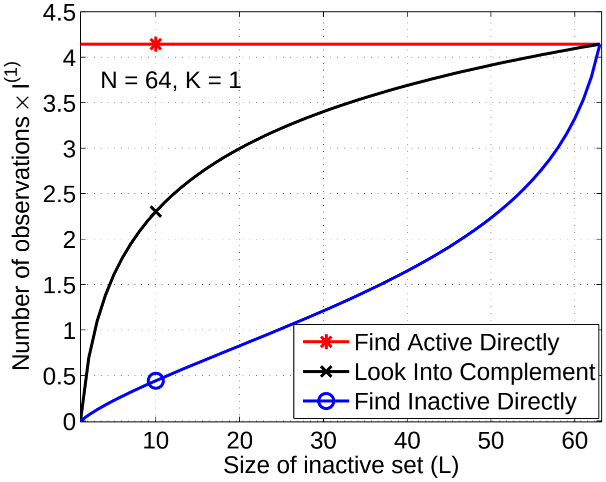

Fig. 2. Sufficiency bounds on the number of observations required to find L inactive variables for $K = 1$ case. The comparison is presented with respect to the value of $MI^{(1)}$, as the application-dependent mutual information term $I^{(1)}$ is common to all the bounds. The approach of finding the L inactive variables directly, especially for small values of L, requires significantly fewer number of observations compared to the approach of finding the inactive variables indirectly, after first identifying the active variables. The plot corresponding to the curve labeled `Find Active Directly` refers to the number of observations that are sufficient for finding the K active variables [19].

The above result is applicable to the abstract signal model specified in Section II. It can be specialized to the non-adaptive group testing model by lower bounding $E_0(\rho, 1, L_0)$, to obtain a relationship between M and the average probability of error for the decoding algorithm. We present the results for the case of the non-adaptive group testing in Section IV.

We first note that $\log \frac{\binom{N}{L+K-1}}{\binom{N-K}{L+K-1}} \approx K \log \left[\frac{N-K}{L_0}\right]$, where L_0 is as defined above, denotes the number of bits that can index all sets (of size $L + K - 1$) for each $L + K - 1$ sized set containing only inactive variables.[5] Thus, the number of observations is approximately $\frac{K}{I^{(K)}} \log \left[\frac{N-K}{L_0}\right]$, where $I^{(K)}$ denotes the mutual information per observation. This explains the first term in (9), as can be seen from the following argument. Using the fact that $\left.\frac{dE_0(\rho, j, n)}{d\rho}\right|_{\rho=0} = nI^{(j)}$, we note that for $\rho = \alpha\frac{K}{L_0}$, $E_0(\rho, 1, L_0) \approx \alpha\frac{K}{L_0}(L_0 I^{(1)})$ for sufficiently small α. Further, in the non-adaptive group testing framework, $I^{(j)} = O(\frac{j}{K})$ [19]. Thus, $E_0(\rho, 1, L_0) \approx \alpha I^{(K)}$. The claim now follows by noting that $\rho \log \binom{N-K}{L_0} \approx \alpha K \log \left[\frac{N-K}{L_0}\right]$. The additional term contributing to the total number of tests in (9) may be an artifact of the particular decoding scheme and/or its analysis presented here.

Before concluding this section and proceeding to specialize the above results to the case of non-adaptive group testing, we summarize a lower bound, derived in [27], on the number of observations required to find a set of L inactive variables. The lower bound will be used in the comparisons and discussion to follow. Let ω denote the index of the defective set such that $S_\omega \subset [N]$ and $|S_\omega| = K$. Given the observation vector, $\underline{\mathbf{y}} \in \mathcal{Y}^M$, let $\phi : \mathcal{Y}^M \times \mathcal{X}^{M \times N} \to \mathcal{S}^H$ denote a decoding function, such that $\hat{S} = \phi(\underline{\mathbf{y}}, \mathbf{X})$ is the decoded set of L inactive variables. Let $P_e = \Pr\left(\left\{\hat{S} \cap S_\omega \neq \{0\}\right\}\right)$. We state a necessary condition on the number of observations in the following theorem.

Theorem 3. *Let N, M, L and K be as defined before. Let $I^{(j)}$ be as defined in (4). A necessary condition on the number of observations M required to find L inactive variables with asymptotically vanishing probability of error, i.e., $\lim_{N\to\infty} P_e = 0$, is given by*

$$M \geq \max_{1 \leq j \leq K} \frac{\Gamma_l(L, N, K, j)}{I^{(j)}}(1 - \eta), \tag{10}$$

where $\Gamma_l(L, N, K, j) \triangleq \log \left[\binom{N-K+j}{j} / \binom{N-K+j-L}{j}\right]$, and for some $\eta > 0$.

The proof is provided in [27]. That is, any sequence of random codebooks that achieves $\lim_{N\to\infty} P_e = 0$ must satisfy the above bound on the length of the codewords. Given a specific application, we can bound $I^{(j)}$ for each $j = 1, 2, \ldots, K$, and obtain a characterization on the necessary number of observations, as we show in the next section.

IV. Finding Non-Defective Items Via Group Testing

In this section, we specialize the above mutual information based results to the case of non-adaptive group testing, and characterize the number of tests to identify a subset of non-defective items in a large population. We consider the random pooling based noisy non-adaptive group testing model given by (2) [3], [19]. Our goal is to find upper bounds on the number of tests required to identify an L sized subset belonging to $[N]\backslash\mathcal{G}$ using the observations $\underline{y}$, with vanishing probability of error as $N \to \infty$. We focus on the regime where $K, L, N \to \infty$ with $\frac{L}{N-K} \to \alpha_0$, $\frac{K}{N-K} \to \beta_0$ for some fixed $\alpha_0, \beta_0 \in (0, 1)$.

First, we make a note about lower bounds on the number of tests. Using the results of Theorem 3, we need to upper bound the mutual information term, $I^{(j)}$, for the group testing signal model given in (2). Using the bounds on $I^{(j)}$ [35], with[6] $p = \frac{1}{K}$ and $u \leq 0.5$, we summarize the order-accurate lower bounds on the number of tests to find a set of L non-defective items in Table I. A brief sketch of the derivation of these results is provided in Appendix VII-B.

To compute the upper bounds on the number of tests, we need to lower bound $E_0(\rho, 1, n)$ for some $\rho \in [0, 1]$ and show that the negative exponent in the probability of error term in (9) can be made strictly greater than 0 by choosing M sufficiently large. We present our lower bounds on $E_0(\rho, 1, n)$ in the following lemma.

[5] Note that the decoding scheme might end up choosing a maximum of $L + K - 1$ inactive variables, and this represents the worst case outcome.

[6] The value of p is a test design parameter. In general, $p = \frac{\alpha}{K}$, with α independent of K, has been widely used in the group testing literature [19]–[21], [35]. In the noiseless case, $p = 1/K$ is a useful choice since it maximizes the mutual information $I^{(j)}$ [19], [35]. In [21], it is shown that choosing $\alpha = \log 2$ helps close the gap between upper and lower bound in the noiseless case.

Lemma 1. *Let N, M, L and K be as defined above. Let $L_0 = (N - L - 2K + 1)$. Let $E_0(\rho, j, n)$ be as defined in (3) and define $\rho_0 \triangleq \frac{K-1}{L_0}$. For the non-adaptive group testing model with $p = \frac{1}{K}$ and for all values of $L \leq (N - 3K + 1)$, we have*

(a) For the noiseless case ($u = 0, q = 0$):

$$E_0(\rho_0, 1, L_0) \geq \frac{(1 - e^{-1}) - (\frac{1}{K})^K}{e}. \tag{11}$$

(b) For the additive noise only case ($u = 0, q > 0$):

$$E_0(\rho_0, 1, L_0) \geq \frac{e^{-2}}{4}(1 - q). \tag{12}$$

(c) For the dilution noise only case ($u > 0, q = 0$):

$$E_0(\rho_0, 1, L_0) \geq \frac{e^{-2}}{4}(1 - u^{\frac{1}{K}}). \tag{13}$$

The proof of the above lemma is presented in Appendix VII-A. For notational convenience, we let $E_0^{(lb)}$ denote a common lower bound on $E_0(\rho_0, 1, L_0)$, as derived above. The following theorem presents an upper bound on the number of tests required to identify L non-defective items in the non-adaptive group testing setup.

Theorem 4. *Let P_e be the average probability of error in finding L inactive variables under the decoding scheme described in Section III-B2. Note that P_e is upper bounded by (9). Let $L_0 \triangleq (N - L - 2K + 1)$ and let $\theta_0 \triangleq \frac{L+K-1}{N-K}$. Then, for any $\epsilon_0 > 0$ and all values of $L \leq (N - 3K + 1)$, if M is chosen as*

$$M > (1 + \epsilon_0)\frac{K-1}{E_0^{(lb)}}\left[\frac{H_b(\theta_0)}{1 - \theta_0} + \log\left(2 + \frac{L}{K-1}\right) + 1 + \frac{\log K}{K-1}\right], \tag{14}$$

then $P_e \leq \exp\left(-\epsilon_0(K-1)\log\frac{N-K}{L_0}\right)$.

An outline of the proof is presented in Section V-C. In the regime where $L, K \to \infty$ as $N \to \infty$, it follows from the above theorem that $\lim_{N\to\infty} P_e = 0$.

Finally, we present an upper bound on the number of tests obtained for the indirect decoding scheme presented in Section III-A for the noiseless case. Using [19, Lemma VII.1 and VII.3] to lower bound $E_0(\rho, j, 1)$ for the noiseless case, and noting that, from the union bound, we have $C_0(j) \leq \frac{j\binom{N-K-1}{L-1}}{\binom{N-K}{L}} = j\frac{L}{N-K}$, the following corollary builds on the result presented in Corollary 1.

Corollary 2. *Let P_e be the average probability of error in finding L inactive variables under the decoding scheme described in Section III-A. Note that P_e is upper bounded by (5). For any $\epsilon_0 > 0$, there exist absolute constants $c_0, c_1 > 2$, independent of N, K and L, such that if M is chosen as*

$$M > (1 + \epsilon_0)c_0 K\left(\log L + c_1 \log^3 K\right), \tag{15}$$

then $P_e \leq \exp\left(-\epsilon_0(K \log L)\right)$.

A. Discussion of the Results

We now make following observations about the results presented in this section.

1) Linear Scaling Regime

First, we consider the linear scaling regime, where, for some fixed $\alpha_0, \beta_0 \in (0, 1)$, $\frac{L}{N-K} \to \alpha_0$, $\frac{K}{N-K} \to \beta_0$ as $N \to \infty$. Since our results apply for $L \leq N - 3K + 1$, we consider α_0, β_0 such that $\alpha_0 + 2\beta_0 \leq 1$. For the direct decoding scheme presented in Section III-B, we summarize the upper bounds on the number of tests to find a set of L non-defective items in Table II.

(a) We first consider the noiseless case.

   (i) For the direct decoding scheme, $O(K)$ number of tests are sufficient. In comparison, using results from Corollary 2, $O(K \log L \log^2 K)$ tests are sufficient for the indirect decoding scheme. Also, from [19, Theorem V.2], $O(K \log N \log^2 K)$ tests are sufficient for finding all the defective items. Thus, in this case, the direct decoding scheme for finding non-defective items performs better compared to the indirect decoding schemes by a poly-log factor of the number of defective items, K. Further, from Table I, we observe that the upper bound on the number of tests for the direct decoding scheme is within a $c \log K$ factor of the lower bound in [27], where c is a constant independent of N, L and K. We thus obtain an $O(\log K)$ improvement over the upper bounds for computationally tractable algorithms such as COMP, COMA, and linear programming based algorithms, where it was shown that the sufficient number of tests required to guarantee non-defective subset recovery are within $O(\log^2 K)$ of the lower bounds [20], [27].

   (ii) The size of non-defective set, L, impacts the upper bound on the number of tests only through α_0, i.e., the fraction of non-defective items that need to be found. From Table II, $\Phi(\alpha_0, \beta_0)$ is an increasing function of α_0. That is, a higher α_0 results in a higher rate at which the upper bound on the number of tests increases with K.

(b) Performance under noisy observations:

   (i) For the additive noise, $O(\frac{K}{1-q})$ number of tests are sufficient for the direct decoding scheme. The indirect scheme (as well as the scheme for finding the defective items) also show similar $\frac{1}{1-q}$ factor increase in the number of tests under additive noise scenario (see, e.g., [19, Theorem VI.2]). Further, from Table I, we observe that, for fixed α_0, β_0 and q, the upper bound on the number of tests for the direct decoding scheme is within a constant factor of the lower bound.

   (ii) For dilution noise, $O\left(\frac{K}{1-u^{\frac{1}{K}}}\right)$ are sufficient for the direct decoding scheme. Another characterization for the sufficient number of tests for the direct decoding scheme, based on the remark at the end

TABLE I

FINDING A SUBSET OF L NON-DEFECTIVE ITEMS: RESULTS FOR NECESSARY NUMBER OF GROUP TESTS WHICH HOLD ASYMPTOTICALLY AS $(N, K, L) \to \infty$, $\frac{L}{N-K} \to \alpha_0$ WITH $0 < \alpha_0 < 1$. THE CONSTANTS $C_n, C'_n, C''_n > 0$ ARE INDEPENDENT OF N, L, K, u AND q.

No Noise	$\frac{C_n K}{\log K} \log \frac{1}{[1-\alpha_0+o(1)]}$
Additive Noise	$\frac{C'_n K}{\min\{\log \frac{1}{q}, \log K\}} \log \frac{1}{[1-\alpha_0+o(1)]}$
Dilution Noise	$\frac{C''_n K}{(1-u)\log K} \log \frac{1}{[1-\alpha_0+o(1)]}$

TABLE II

FINDING A SUBSET OF L NON-DEFECTIVE ITEMS: RESULTS FOR SUFFICIENT NUMBER OF GROUP TESTS WHICH HOLD ASYMPTOTICALLY AS $(N, K, L) \to \infty$, $\frac{L}{N-K} \to \alpha_0$ AND $\frac{K}{N-K} \to \beta_0$ WITH $0 < \beta_0, \alpha_0 < 1$ SUCH THAT $\alpha_0 + 2\beta_0 < 1$. DEFINE $\Phi(\alpha_0, \beta_0) \triangleq \left(\frac{H_b(\gamma_0)}{1-\gamma_0} + \log(2 + \frac{\alpha_0}{\beta_0}) + 1\right)$, WHERE $\gamma_0 = \alpha_0 + \beta_0$. THE CONSTANTS $C_s, C'_s, C''_s > 0$ ARE INDEPENDENT OF N, L, K, u AND q.

No Noise	$\frac{C_s K}{(1-o(1))} [\Phi(\alpha_0, \beta_0) + o(1)]$
Additive Noise	$\frac{C'_s K}{(1-q)} [\Phi(\alpha_0, \beta_0) + o(1)]$
Dilution Noise	$\frac{C''_s K}{(1-u^{\frac{1}{K}})} [\Phi(\alpha_0, \beta_0) + o(1)]$

of Appendix VII-A, is $O\left(\frac{K^2}{1-u^{\frac{1}{2}}}\right)$ number of tests. The direct decoding scheme shows high sensitivity to the dilution noise. This behavior is in sharp contrast to the indirect scheme, where the dilution noise parameter u leads to an increase in the number of tests only by a factor of $\frac{c}{1-u}$ (see, e.g., [19, Theorem VI.5]). From Table I, the lower bounds also show an increase in the number of tests by a factor $\frac{1}{1-u}$ for the dilution noise scenario. The conservativeness of the upper bound for the direct decoding scheme in the presence of dilution noise may be due to: (a) The lower bound on E_0 is $\Omega(\frac{1}{K})$, which underscores the general fact that the group testing system is more sensitive to the diluton noise, and (b) The term $\log\binom{N-1-L_0}{K-1}$ in (9), which might be due to the particular decoding scheme employed or the specific technique employed in bounding the error exponent.

2) *Sub-Linear Scaling Regime*

We now consider the sub-linear scaling regime, where $\frac{L}{N} \to 0$, $\frac{K}{N} \to 0$ as $N \to \infty$. In particular, we consider $L = N^{\lambda}$ and $K = N^{\lambda'}$ with $0 < \lambda, \lambda' < 1$. We discuss the noiseless case; similar conclusions can be drawn under noisy observations. The lower bound scales as $o(1)$, since, in this regime, $\frac{L}{N-K} \to 0$. However, there are two contrasting scenarios for the upper bounds. When $\lambda < \lambda'$, $O(K)$ tests are sufficient. Further, compared to the linear regime, we obtain smaller constants, since both the terms $\frac{H_b(\theta_0)}{1-\theta_0}$ and $\frac{L}{K-1}$ in (14) vanish asymptotically as $N \to \infty$. Also, the direct decoding scheme offers significant gains compared to the indirect decoding scheme, since the necessary number of tests required for finding defective items scales as $\Omega(K \log N)$. Thus, we again obtain an improvement of $O(\log N)$ over computationally tractable algorithms [20], [27]. However, when $\lambda > \lambda'$, $O(K \log N^{\alpha})$ tests are sufficient, where α $(\triangleq \lambda - \lambda') < 1$. This is because of the $\log\left(2 + \frac{L}{K-1}\right)$ term in (14), which scales as $O(\log N^{\alpha})$. In this case, we obtain an improvement in the constant involved, with the gain depending on the the difference between λ and λ'. This regime also exposes a non-trivial gap between the upper and lower bounds, indicating the need for further work into finding better decoding schemes or tighter bounds to close the gap.

V. PROOFS OF THE MAIN RESULTS

We now present the proofs of Theorems 1, 2 and 4, which are the main results in this paper.

A. Proof of Theorem 1: Sufficient Number of Observations, $K = 1$

At the heart of the proof of this theorem is the derivation of an upper bound on the average probability of error in finding L inactive variables using the decoding scheme described in Section III-B1. This is obtained by characterizing the exponents on the average probability of error [29]. Without loss of generality, due to the symmetry in the model, we can assume that X_1 is active. Then, the decoding algorithm makes an error if $P(\underline{\mathbf{y}}|\mathbf{X}_1)$ falls within the last L entries of the sorted array generated as described in the decoding scheme. Let $\underline{\mathbf{y}}$ be the observed output, and let $\mathcal{E}$ denote the event that an error has occurred, when item X_1 is the active variable and $\mathbf{X}_1$ is the first column of $\mathbf{X}$. Further, let $\Pr(\mathcal{E})$ be a shorthand for $\Pr\{\text{error}|X_1 \text{ is active}, \mathbf{X}_1, \underline{\mathbf{y}}\}$. The overall average probability of error, P_e, can be expressed as

$$P_e = \sum_{\underline{\mathbf{y}}, \mathbf{X}_1} P(\underline{\mathbf{y}}|\mathbf{X}_1) Q(\mathbf{X}_1) \Pr(\mathcal{E}). \tag{16}$$

Let $S_z \subset [N] \backslash 1$ be a set of $N-L$ items, i.e., $|S_z| = N-L$. Let $\mathcal{S}_{\mathbf{z}}$ denote the set of all possible S_z. Further, let $\mathcal{A}_{S_z} \subset \{\mathbf{X}_{S_z}\}$ be such that, $\mathcal{A}_{S_z} = \{\{\mathbf{X}_{S_z}\} : P(\underline{\mathbf{y}}|\mathbf{X}_j) \geq P(\underline{\mathbf{y}}|\mathbf{X}_1) \ \forall\ j \in S_z\}$. That is, $\mathcal{A}_{S_z}$ represents all those realizations of the random variable $\mathbf{X}_{S_z}$ which satisfy the above condition, which states that each variable in S_z is more likely than the active variable, X_1. It is easy to see that $\mathcal{E} \subset \mathcal{A} \triangleq \bigcup_{S_z \in \mathcal{S}_z} \mathcal{A}_{S_z}$, i.e., an error event implies that there exists at least one set of $N-L$ variables, S_z, such that $P(\underline{\mathbf{y}}|\mathbf{X}_j) \geq P(\underline{\mathbf{y}}|\mathbf{X}_1) \ \forall\ j \in S_z$. Thus, $\Pr(\mathcal{E}) \leq \Pr(\mathcal{A})$. Let s be an optimization variable such that $0 \leq s \leq 1$. The following set of inequalities upper bound $\Pr(\mathcal{E})$:

$$\begin{aligned}
\Pr(\mathcal{E}) &\leq \sum_{S_z \in \mathcal{S}_{\mathbf{z}}} \sum_{\mathbf{X}_{S_z} \in \mathcal{A}_{S_z}} Q(\mathbf{X}_{S_z}) \\
&\overset{(a)}{\leq} \sum_{S_z \in \mathcal{S}_{\mathbf{z}}} \sum_{\mathbf{X}_{S_z} \in \mathcal{A}_{S_z}} Q(\mathbf{X}_{S_z}) \prod_{j \in S_z} \left[\frac{P(\underline{\mathbf{y}}|\mathbf{X}_j)}{P(\underline{\mathbf{y}}|\mathbf{X}_1)}\right]^s \\
&\overset{(b)}{\leq} \sum_{S_z \in \mathcal{S}_{\mathbf{z}}} \sum_{\mathbf{X}_{S_z}} \prod_{j \in S_z} Q(\mathbf{X}_j) \left[\frac{P(\underline{\mathbf{y}}|\mathbf{X}_j)}{P(\underline{\mathbf{y}}|\mathbf{X}_1)}\right]^s \\
&\overset{(c)}{=} \sum_{S_z \in \mathcal{S}_{\mathbf{z}}} \prod_{j \in S_z} \sum_{\mathbf{X}_j} Q(\mathbf{X}_j) \left[\frac{P(\underline{\mathbf{y}}|\mathbf{X}_j)}{P(\underline{\mathbf{y}}|\mathbf{X}_1)}\right]^s
\end{aligned}$$

$$\overset{(d)}{=} \binom{N-1}{L-1} \left\{ \sum_{\mathbf{X}_j} Q(\mathbf{X}_j) \left[\frac{P(\underline{\mathbf{y}}|\mathbf{X}_j)}{P(\underline{\mathbf{y}}|\mathbf{X}_1)} \right]^s \right\}^{N-L}. \quad (17)$$

In the above, (a) follows since we are multiplying with terms that are all greater than 1 and (b) follows since we are adding extra nonnegative terms by summing over all $\mathbf{X}_{S_z}$. (c) follows by using the independence of the codewords, i.e., $Q(\mathbf{X}_{S_z}) = \prod_{j \in S_z} Q(\mathbf{X}_j)$, and simplifying. (d) follows since the value of the expression inside the product term does not depend upon any particular j.

Let $0 \leq \rho \leq 1$. If the R.H.S. in (17) is less than 1, then raising it to the power ρ makes it bigger, and if it is greater than 1, it remains greater than 1 after raising it to the power ρ. Thus, we get the following upper bound on Pr$(\mathcal{E})$:[7]

$$\Pr(\mathcal{E}) \leq \binom{N-1}{L-1}^{\rho} \left\{ \sum_{\mathbf{X}_j} Q(\mathbf{X}_j) \left[\frac{P(\underline{\mathbf{y}}|\mathbf{X}_j)}{P(\underline{\mathbf{y}}|\mathbf{X}_1)} \right]^s \right\}^{\rho(N-L)}. \quad (18)$$

Substituting this into (16) and simplifying, we get

$$P_e \leq \binom{N-1}{L-1}^{\rho} \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_1} Q(\mathbf{X}_1) P(\underline{\mathbf{y}}|\mathbf{X}_1)^{1-\rho(N-L)s} \times \left\{ \sum_{\mathbf{X}_j} Q(\mathbf{X}_j) P(\underline{\mathbf{y}}|\mathbf{X}_j)^s \right\}^{\rho(N-L)}. \quad (19)$$

Putting $s = 1/(1+\rho(N-L))$, we get

$$P_e \leq \binom{N-1}{L-1}^{\rho} \sum_{\underline{\mathbf{y}}} \left\{ \sum_{\mathbf{X}_j} Q(\mathbf{X}_j) P(\underline{\mathbf{y}}|\mathbf{X}_j)^{\frac{1}{1+\rho(N-L)}} \right\}^{1+\rho(N-L)}. \quad (20)$$

Finally, using the independence across observations and using the definition of $E_0(\rho, j, n)$ from (3) with $j = 1$ and $n = N-L$, we get

$$\begin{aligned} P_e &\leq \binom{N-1}{L-1}^{\rho} \\ &\left[\sum_{Y \in \mathcal{Y}} \left\{ \sum_{X_j \in \mathcal{X}} Q(X_j) P(Y|X_j)^{\frac{1}{1+\rho(N-L)}} \right\}^{1+\rho(N-L)} \right]^M \\ &= \exp^{[-MF(\rho)]}, \end{aligned} \quad (21)$$

where $F(\rho) = E_0(\rho, 1, N-L) - \frac{\rho \log \binom{N-1}{L-1}}{M}$. Hence (6) follows.

For the following discussion, we treat F and E_0 as functions of ρ only and all the derivatives are with respect to ρ. Note that $F^{'}(\rho) = E_0^{'}(\rho) - \frac{\log \binom{N-1}{L-1}}{M}$. It is easy to see that $E_0(0) = 0$ and hence $F(0) = 0$. With some calculation, we get,

$$\begin{aligned} E_0^{'}(\rho)\Big|_{\rho=0} &= (N-L) \sum_{Y,X} P(Y,X) \log \frac{P(Y|X)}{P(Y)} \\ &= (N-L) I^{(1)}. \end{aligned} \quad (22)$$

Using the Taylor series expansion of $E_0(\rho)$, and following similar analysis as in [19, Section III.D], it is easy to show that there exists a $\rho \in (0,1]$, sufficiently small, such that if M is chosen as in (7), then $MF(\rho) > \epsilon_1 (N-L) \log \binom{N-1}{L-1}$ for some $\epsilon_1 > 0$, independent of N and L. This completes the proof.

Remark: For the decoding scheme described in Section III-A, for the $K = 1$ case, using similar arguments as the above, it is easy to show that if $M > (1+\epsilon_0) \frac{\log L}{I^{(1)}}$ for any $\epsilon_0 > 0$, then there exists $\epsilon_1 > 0$, and independent of N and L, such that $P_e \leq \exp(-\epsilon_1 \log L)$, i.e., $P_e \to 0$, as $L \to \infty$.

B. Proof of Theorem 2: Sufficient Number of Observations, $K > 1$

The decoding algorithm outputs a set, S_H, of at least L inactive variables. A decoding error happens if the set S_H contains one or more variables from the active set. We now upper bound the average probability of error of the proposed decoding algorithm. The probability is averaged over all possible instantiations of $\{\mathbf{X}, \underline{\mathbf{y}}\}$ as well as over all possible active sets. By symmetry of the codebook ($\mathbf{X}$) construction, the average probability of error is the same for all the active sets. Hence, we fix the active set and then compute average probability of error with this set. Let $S_1 \subset [N]$ be the active set such that $|S_1| = K$. We also define the following notation: For any set $S_\omega \subset [N]$ such that $|S_\omega| = K$ and for any item $j \in S_\omega$, let $S_{\omega j^c} \triangleq S_\omega \backslash j$. Note that $|S_{\omega j^c}| = K-1$.

For any $d \in S_1$, let $\mathcal{E}_d$ be the error event $d \in S_H$. The overall average probability of error, P_e, in finding L inactive variables can then be upper bounded as

$$P_e \leq \sum_{d \in S_1} \Pr(\mathcal{E}_d). \quad (23)$$

Further,

$$\begin{aligned} \Pr(\mathcal{E}_d) = \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_{S_1}} & P(\underline{\mathbf{y}}|\mathbf{X}_{S_1}) Q(\mathbf{X}_{S_1}) \\ & \left[\Pr\{\mathcal{E}_d | S_1 \text{ is the active set}, \underline{\mathbf{y}}, \mathbf{X}_{S_1}\} \right]. \end{aligned} \quad (24)$$

We now upper bound $\Pr\{\mathcal{E}_d | S_1 \text{ is the active set}, \underline{\mathbf{y}}, \mathbf{X}_{S_1}\}$. Let $S_z \subset [N] \backslash S_1$ be such that $|S_z| = L_0$. Let $S_\omega \subset [N]$ be a K sized index set such that $S_\omega = \{d \cup S_{\omega d^c}\}$, where $S_{\omega d^c} \subset [N] \backslash \{d\} \backslash S_z$ and $d \in S_1$ (see Figure 3). Further, let $\mathcal{S}_{\mathbf{z}}$ and $\mathcal{S}_{\omega d^c}$ be the collection of all possible sets S_z and $S_{\omega d^c}$, respectively. It is easy to see that $|\mathcal{S}_{\mathbf{z}}| = \binom{N-K}{L_0}$ and $|\mathcal{S}_{\omega d^c}| = \binom{N-1-L_0}{K-1}$. With S_1 as the active set, $d \in S_1$, the observed output $\underline{\mathbf{y}}$ and the codebook entries corresponding to set S_1 as $\mathbf{X}_{S_1}$, define $\mathcal{A}_d(S_z, S_{\omega d^c}) \subset \{\mathbf{X}_{S_z \cup S_{\omega d^c}}\}$ and $\mathcal{A}_d$ as follows:

$$\begin{aligned} \mathcal{A}_d(S_z, S_{\omega d^c}) = \{ & \{\mathbf{X}_{S_z}, \mathbf{X}_{S_{\omega d^c}}\} : \\ & P(\underline{\mathbf{y}}|\mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}}) \geq P(\underline{\mathbf{y}}|\mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \ \forall \ \alpha \in S_z \}, \end{aligned} \quad (25)$$

$$\mathcal{A}_d = \bigcup_{S_z \in \mathcal{S}_z} \bigcup_{S_{\omega d^c} \in \mathcal{S}_{\omega d^c}} \mathcal{A}_d(S_z, S_{\omega d^c}). \quad (26)$$

[7] This is a standard Gallager bounding technique [29, Section 5.6].

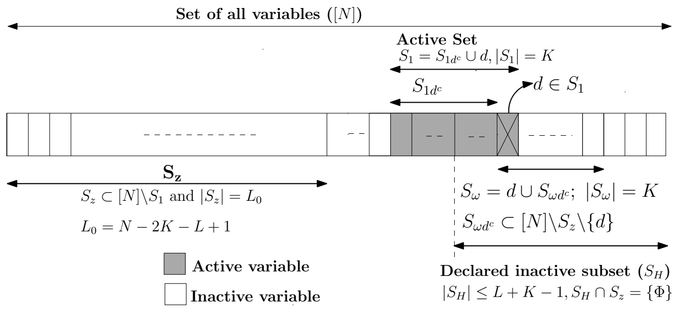
Set of all variables ($[N]$)
Active Set
$S_1 = S_{1d^c} \cup d, |S_1| = K$
S_{1d^c}
$d \in S_1$
$\mathbf{S_z}$
$S_z \subset [N] \backslash S_1$ and $|S_z| = L_0$
$L_0 = N - 2K - L + 1$
$S_\omega = d \cup S_{\omega d^c}$; $|S_\omega| = K$
$S_{\omega d^c} \subset [N] \backslash S_z \backslash \{d\}$
Active variable
Inactive variable
Declared inactive subset (S_H)
$|S_H| \leq L + K - 1, S_H \cap S_z = \{\Phi\}$

Fig. 3. Illustration of the notation used in the proof of Theorem 2.

That is, $\mathcal{A}_d(S_z, S_{\omega d^c})$ represents the set of those realizations of the random variables $\mathbf{X}_{S_z}$ and $\mathbf{X}_{S_{\omega d^c}}$ which satisfy the condition in (25).

Proposition 1. *Pr*$\{\mathcal{E}_d | S_1$ *is the active set*$, \underline{\mathbf{y}}, \mathbf{X}_{S_1}\} \leq$ *Pr*$(\mathcal{A}_d)$

Proof: We will show that given the active set S_1, $d \in S_1$, $\underline{\mathbf{y}}$ and $\mathbf{X}_{S_1}$, the event $\{d \in S_H\}$, i.e., the decoded set of inactive variables contains d, implies the event $\mathcal{A}_d$. We first note that, since $|S_H| \leq L + K - 1$, there exists a set of $L_0 = N - K - (L + K - 1)$ inactive variables that do not belong to S_H. Let $S_z \subset [N] \backslash S_1$ be such a set of inactive variables such that $|S_z| = L_0$ and $S_z \cap S_H = \{\emptyset\}$.

Further, since $d \in S_H$, this implies that there exits an $\omega \in \Omega_{\text{last}}$ such that d belongs to S_ω, where Ω_{last} is as defined in the decoding scheme for $K > 1$ (see Section III-B2). With the notation described above, we can represent such S_ω as $\{d \cup S_{\omega d^c}\}$, where $S_{\omega d^c} \subset [N] \backslash \{d\} \backslash S_z$ such that $|S_{\omega d^c}| = K - 1$. For any $\alpha \in S_z$, if we replace $d \in S_\omega$ with α and evaluate $P(\underline{\mathbf{y}} | \mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}})$, it cannot be smaller than $P(\underline{\mathbf{y}} | \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})$ or else the decoding algorithm would have chosen α as belonging to S_H. This implies that, there exists a realization of $\mathbf{X}_{S_z}$ and $\mathbf{X}_{S_{\omega d^c}}$ such that $P(\underline{\mathbf{y}} | \mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}}) \geq P(\underline{\mathbf{y}} | \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})$ $\forall$ $\alpha \in S_z$, i.e., $\mathcal{A}_d$ occurs. ■

We now upper bound $\Pr(\mathcal{A}_d)$ as follows:

$$\Pr(\mathcal{A}_d) \leq \sum_{S_z \in \mathcal{S}_\mathbf{z}} \sum_{S_{\omega d^c} \in \mathcal{S}_{\omega \mathbf{d}^\mathbf{c}}} q_d, \tag{27}$$

where $q_d \triangleq \Pr\{\mathcal{A}_d(S_z, S_{\omega d^c}) | S_1$ is active set$, \underline{\mathbf{y}}, \mathbf{X}_{S_1}\}$. Here, the randomness comes from the set of variables in S_z and $S_{\omega d^c}$, i.e., $\mathbf{X}_{S_z}$ and $\mathbf{X}_{S_{\omega d^c}}$. Let s be such that $0 \leq s \leq 1$. We have

$$\begin{aligned}
q_d = & \sum_{\mathbf{X}_{S_z}, \mathbf{X}_{S_{\omega d^c}} \in \mathcal{A}_d(S_z, S_{\omega d^c})} Q(\mathbf{X}_{S_z}, \mathbf{X}_{S_{\omega d^c}}) \\
\overset{\text{(a)}}{\leq} & \sum_{\mathbf{X}_{S_{\omega d^c}}, \mathbf{X}_{S_z} \in \mathcal{A}_d(S_z, S_{\omega d^c})} Q(\mathbf{X}_{S_z}, \mathbf{X}_{S_{\omega d^c}}) \\
& \times \prod_{S_\alpha \in S_z} \left[\frac{P(\underline{\mathbf{y}} | \mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}})}{P(\underline{\mathbf{y}} | \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})} \right]^s \\
\overset{\text{(b)}}{\leq} & \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \sum_{\mathbf{X}_{S_z}} Q(\mathbf{X}_{S_z}) \\
& \times \prod_{S_\alpha \in S_z} \left[\frac{P(\underline{\mathbf{y}} | \mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}})}{P(\underline{\mathbf{y}} | \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})} \right]^s \\
\overset{\text{(c)}}{=} & \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \prod_{l=1}^{L_0} \sum_{\mathbf{X}_{S_\alpha}} Q(\mathbf{X}_{S_\alpha}) \left[\frac{P(\underline{\mathbf{y}} | \mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}})}{P(\underline{\mathbf{y}} | \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})} \right]^s \\
\overset{\text{(d)}}{=} & \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \left\{ \sum_{\mathbf{X}_{S_\alpha}} Q(\mathbf{X}_{S_\alpha}) \left[\frac{P(\underline{\mathbf{y}} | \mathbf{X}_\alpha, \mathbf{X}_{S_{\omega d^c}})}{P(\underline{\mathbf{y}} | \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})} \right]^s \right\}^{L_0} \\
= & \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \underbrace{\left\{ \sum_{\mathbf{X}_{S_\alpha}} Q(\mathbf{X}_{S_\alpha}) \left[\frac{P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_\alpha)}{P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_d)} \right]^s \right\}^{L_0}}_{\triangleq \mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})}.
\end{aligned} \tag{28}$$

In the above, (a)-(d) follow using the same reasoning as in (17) in the proof of Theorem 1 (Section V-A). We note that, due to symmetry in the construction of codebook, $\mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}})$ does not depend upon the index set S_z or $\mathbf{X}_{S_z}$. In fact, it depends only upon the given realizations of $\mathbf{X}_{S_{\omega d^c}}$, $\mathbf{X}_d$ and not on the particular index sets $S_{\omega d^c}$ and d, respectively. Thus, from (27), and for some $0 \leq \rho \leq 1$, we get

$$\begin{aligned}
\Pr(\mathcal{A}_d) \leq & \sum_{S_{\omega d^c} \in \mathcal{S}_{\omega \mathbf{d}^\mathbf{c}}} \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \left[\sum_{S_z \in \mathcal{S}_z} \mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \right] \\
\leq & \sum_{S_{\omega d^c} \in \mathcal{S}_{\omega \mathbf{d}^\mathbf{c}}} \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \\
& \times \left[\sum_{S_z \in \mathcal{S}_z} \mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \right]^\rho \\
\leq & \binom{N - 1 - L_0}{K - 1} \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \\
& \times \left[\binom{N - K}{L_0} \mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \right]^\rho.
\end{aligned} \tag{29}$$

The second inequality above follows since the expression inside the square brackets represents the probability of a union of events and therefore, as in the $K = 1$ case, by raising it to a power $0 < \rho \leq 1$, we still get an upper bound [29, Section 5.6]. Let $C_2 \triangleq \binom{N-K}{L_0}^\rho \binom{N-1-L_0}{K-1}$. Using proposition 1, we substitute the above expression into (23) to get:

$$\begin{aligned}
\Pr(\mathcal{E}_d) \leq \; & C_2 \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_{S_1}} Q(\mathbf{X}_{S_1}) P(\underline{\mathbf{y}} | \mathbf{X}_{S_1}) \\
& \times \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_{\omega d^c}}) \left[\mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \right]^\rho \\
\overset{\text{(a)}}{\leq} \; & C_2 \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_{S_1}} \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_{S_1}) P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_{S_1}) \\
& \times \left[\mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \right]^\rho \\
\overset{\text{(b)}}{\leq} \; & C_2 \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_d} \sum_{\mathbf{X}_{1d^c}} \sum_{\mathbf{X}_{S_{\omega d^c}}} Q(\mathbf{X}_d) \\
& \times P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}}, \mathbf{X}_{S_{1d^c}} | \mathbf{X}_d) \left[\mathcal{P}_0(\underline{\mathbf{y}}, \mathbf{X}_d, \mathbf{X}_{S_{\omega d^c}}) \right]^\rho \\
\overset{\text{(c)}}{\leq} \; & C_2 \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_{S_{\omega d^c}}} \sum_{\mathbf{X}_d} Q(\mathbf{X}_d) \; P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_d)
\end{aligned}$$

$$
\times \left\{ \sum_{\mathbf{X}_{S_\alpha}} Q(\mathbf{X}_{S_\alpha}) \left[\frac{P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_\alpha)}{P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_d)} \right]^s \right\}^{\rho L_0}
$$

$$
\overset{\text{(d)}}{\leq} C_2 \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_{S_{\omega d^c}}} \left\{ \sum_{\mathbf{X}_{S_\alpha}} Q(\mathbf{X}_{S_\alpha}) \right.
$$

$$
\left. \times P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_\alpha)^{\frac{1}{1+\rho L_0}} \right\}^{1+\rho L_0} . \tag{30}
$$

In the above equation, (a) follows because given the active set S_1, $\underline{\mathbf{y}}$ is independent of the other input variables. Thus, $P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_{S_1}) = P(\underline{\mathbf{y}} | \mathbf{X}_{S_1}) Q(\mathbf{X}_{S_{\omega d^c}})$. (b) follows since $S_1 = \{d \cup S_{1d^c}\}$. (c) follows by substituting the expression for $\mathcal{P}_0$ and by averaging out $\mathbf{X}_{S_{1d^c}}$, since the expression for $\mathcal{P}_0$ does not depend upon $\mathbf{X}_{S_{1d^c}}$. In (c), the term $[P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_d)]^{s\rho L_0}$ can be factored out from expression inside the curly braces. Finally, (d) is obtained by choosing $s = \frac{1}{1+\rho L_0}$ and simplifying. Next, the above upper bound for $\Pr(\mathcal{E}_d)$ depends only on $\mathbf{X}_d$ and not on any particular value of d. Thus, from (23) and (30) we get:

$$
P_e \leq K C_2 \sum_{\underline{\mathbf{y}}} \sum_{\mathbf{X}_{S_{\omega d^c}}} \left\{ \sum_{\mathbf{X}_{S_\alpha}} Q(\mathbf{X}_{S_\alpha}) P(\underline{\mathbf{y}}, \mathbf{X}_{S_{\omega d^c}} | \mathbf{X}_\alpha)^{\frac{1}{1+\rho L_0}} \right\}^{1+\rho L_0}
$$

$$
\leq \exp\left[-M \left(E_0(\rho, 1, L_0) - \frac{\log(K C_2)}{M} \right) \right]. \tag{31}
$$

The inequality above is obtained by further simplifying using independence across different observations and writing the bound in the exponential form, as in the $K = 1$ case. The upper bound on P_e given in (9) now follows by substituting the value of C_2 in the above. Hence the proof follows. ■

C. Proof of Theorem 4

In (9), consider the term $\mathcal{T}(\rho) \triangleq \left(M E_0(\rho, 1, L_0) - \rho \log \binom{N-K}{L_0} - \log \left[K \binom{N-1-L_0}{K-1} \right] \right)$. Using the results of Lemma 1, for any $\epsilon_0 > 0$, at $\rho = \rho_0$ where $\rho_0 = \frac{K-1}{L_0}$,[8] if M is chosen as

$$
M > (1 + \epsilon_0) \left[\frac{\rho_0 \log \binom{N-K}{L_0}}{E_0^{(lb)}} + \frac{\log \left[\binom{L+2(K-1)}{K-1} \right]}{E_0^{(lb)}} + \frac{\log K}{E_0^{(lb)}} \right], \tag{32}
$$

then, $\mathcal{T}(\rho) > \epsilon_0 (K-1) \left(\log \frac{N-K}{L_0} + \log(2 + \frac{L}{K-1}) \right) > \epsilon_0 (K-1) \log \frac{N-K}{L_0} > 0$.

Using Stirling's formula, for any $n \in \mathbb{Z}_+$: $\sqrt{2\pi} n^{n+1/2} e^{-n} \leq n! \leq e n^{n+1/2} e^{-n}$, we note

$$
\log \binom{N-K}{L_0} \leq L_0 \log(\frac{N-K}{L_0}) + \frac{1}{2} \log \frac{N-K}{L_0(L+K-1)}
$$

$$
+ (L+K-1) \log(\frac{N-K}{L+K-1}) \tag{33}
$$

$$
\leq L_0 \log(\frac{N-K}{L_0})
$$

$$
+ (L+K-1) \log(\frac{N-K}{L+K-1}). \tag{34}
$$

The second inequality follows since under the assumptions on the range of L, $\frac{N-K}{L_0(L+K-1)} < 1$. Thus, with $\theta_0 \triangleq \frac{L+K-1}{N-K}$, we get $\frac{\log \binom{N-K}{L_0}}{L_0} \leq \frac{H_b(\theta_0)}{1-\theta_0}$. Finally, the bound in (14) results by using the inequality $\binom{m}{n} \leq \left(\frac{em}{n}\right)^n$ to upper bound the second term in (32).

VI. Conclusions

In this paper, we considered the problem of identifying L non-defective items out of a large population of N items containing K defective items in a general sparse signal modeling setup. We contrasted two approaches: identifying the defective items using the observations followed by picking L items from the complement set, and directly identifying non-defective items from the observations. We derived upper bounds on the number of observations required for identifying the L non-defective items. We showed that a gain in the number of observations is obtainable by directly identifying the non-defective items. We specialized the results to the nonadaptive group testing setup. We also characterized the number of tests that are sufficient to identify a subset of non-defective items in a large population, under both dilution and additive noise models. Future work could focus on tightening the upper bounds on the sufficient number of tests and/or devising better decoding algorithms, and obtaining order-optimal results.

VII. Appendix

A. Proof of Lemma 1

From (3), it follows that:

$$
E_0(\rho, j, n) = -\log \sum_{Y \in \mathcal{Y}, X_{S(K-j)} \in \mathcal{X}^{K-j}} Q\left(X_{S(K-j)}\right)
$$

$$
\times \left\{ \sum_{X_{S(j)} \in \mathcal{X}^j} Q(X_{S(j)}) \left(P(Y | X_{S(K-j)}, X_{S(j)})\right)^{\frac{1}{1+\rho n}} \right\}^{1+\rho n} \tag{35}
$$

In the above, we substitute $j = 1$, $n = L_0$ and $\rho = \rho_0$. Let $w(X_{S(K-1)})$ denote the number of 1's in $X_{S(K-1)} \in \{0,1\}^{(K-1)}$. Let $n_0 \triangleq 1 + \rho_0 L_0$ and further, note that $n_0 = K$. For the non-adaptive group testing signal model, using (2), we have computed the posterior probability $P(Y | X_{S(K-1)}, X_{S(1)})$ for different scenarios and summarized it in Table III.

(a) Noiseless case: Using $q = 0$, $u = 0$ in Table III and substituting in (35) we get (also see Figure 4):

$$
E_0(\rho, 1, L_0) = -\log \left[1 - (1-p)^{(K-1)} \right.
$$

$$
\left. \times (1 - (1-p)^{n_0} - p^{n_0}) \right]. \tag{36}
$$

Using, (i) the inequality $-\log(1-x) \geq x$ for $x < 1$, (ii) For $p = \frac{1}{K}$, $(1-p)^{(K-1)} > e^{-1}$ and $(1-p)^K < e^{-1}$, (11) results.

(b) Additive noise case: Using $u = 0$ in Table III and substituting in (35) we get:

$$
E_0(\rho, 1, L_0) = -\log \left[1 - (1-p)^{(K-1)} \right.
$$

[8] Note that, for $L \leq N - 3K + 1$, $\rho_0 = \frac{K-1}{L_0} < 1$.

TABLE III
$P(Y|X_{S(K-1)}, X_{S(1)})$ FOR THE NON-ADAPTIVE GROUP TESTING MODEL, UNDER DIFFERENT SCENARIOS.

	$w(X_{S(K-1)})=0$		$w(X_{S(K-1)})=l, 1\leq l\leq K-1$	
	$X_{S(1)}=0$	$X_{S(1)}=1$	$X_{S(1)}=0$	$X_{S(1)}=1$
$P(Y=0\|X_{S(K-1)}, X_{S(1)})$	$(1-q)$	$(1-q)u$	$(1-q)u^l$	$(1-q)u^{l+1}$
$P(Y=1\|X_{S(K-1)}, X_{S(1)})$	q	$(1-(1-q)u)$	$1-(1-q)u^l$	$1-(1-q)u^{l+1}$

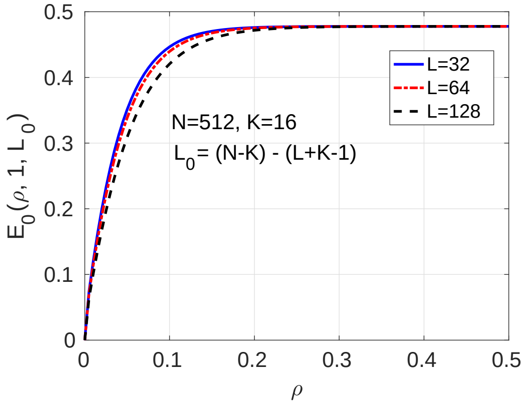

Fig. 4. $E_0(\rho, 1, L_0)$ vs ρ: Noiseless case.

$$\times\left(1-(1-q)(1-p)^{n_0}-\left\{(1-p)q^{\frac{1}{n_0}}+p\right\}^{n_0}\right)\Bigg]. \tag{37}$$

To lower bound E_0, we first upper bound the term $t_0 \triangleq \left\{(1-p)q^{\frac{1}{n_0}}+p\right\}^{n_0}$. For any $n\geq 1$, x^n is a convex function and hence, using Jensen's inequality we get $t_0\leq(1-p)q+p$. Substituting and further simplifying we get:

$$\begin{aligned}E_0(\rho,1,L_0)\geq -\log\big[1-(1-p)^K(1-q)\\ \times\left(1-(1-p)^{(n_0-1)}\right)\big].\end{aligned} \tag{38}$$

The bound in (12) now results by using the inequality $-\log(1-x)\geq x$ for $x<1$ and noting the following: For $p=\frac{1}{K}$, using the inequality, $1-x\leq e^{-x}\leq 1-\frac{x}{2}$ for $0\leq x\leq 1$, we get $(1-p)^K\geq e^{-2}$ and $1-(1-p)^{(n_0-1)}\geq \frac{n_0-1}{2K}\geq\frac{1}{4}$ for $K\geq 2$.

(c) Dilution noise case: Let $G_l \triangleq \binom{K-1}{l}p^l(1-p)^{(K-1-l)}$. Using $q=0$ in Table III and substituting in (35) we get:

$$E_0(\rho,1,L_0) = -\log\left[T_0+T_1\right], \tag{39}$$

where $T_0\triangleq\sum_{l=0}^{K-1}G_l u^l\left((1-p)+pu^{\frac{1}{n_0}}\right)^{n_0}$ and $T_1\triangleq\sum_{l=0}^{K-1}G_l\Big((1-p)(1-u^l)^{\frac{1}{n_0}}+p(1-u^{l+1})^{\frac{1}{n_0}}\Big)^{n_0}$. Using Jensen's inequality to upper bound T_1, we get

$$T_1\leq\sum_{l=0}^{K-1}G_l\left((1-p)(1-u^l)+p(1-u^{l+1})\right) \tag{40}$$

$$=1-\zeta_0\sum_{l=0}^{K-1}G_l u^l, \tag{41}$$

where $\zeta_0\triangleq(1-(1-u)p)$ and we have made use of the fact that $\sum_{l=0}^{K-1}G_l=1$. Further, since $\sum_{l=0}^{K-1}G_l u^l=\zeta_0^{(K-1)}$, we get

$$E_0(\rho,1,L_0)\geq-\log\left[1-(\zeta_0-\psi_0)\zeta_0^{(K-1)}\right], \tag{42}$$

where $\psi_0\triangleq\left(1-(1-u^{\frac{1}{n_0}})p\right)^{n_0}$. Using the inequality $-\log(1-x)\geq x$ for $x<1$, we get:

$$\begin{aligned}E_0(\rho,1,L_0)&\geq(\zeta_0-\psi_0)\zeta_0^{(K-1)}\\ &\geq\left[1-\left(1-(1-u^{\frac{1}{n_0}})p\right)^{n_0-1}\right]\zeta_0^K,\end{aligned} \tag{43}$$

where the second inequality follows since $(1-(1-u^{\frac{1}{n_0}})p)\geq\zeta_0$. The bound in (13) now results by noting the following: For $p=\frac{1}{K}$, using the inequality, $1-x\leq e^{-x}\leq 1-\frac{x}{2}$ for $0\leq x\leq 1$, we get $\zeta_0^K\geq e^{-2(1-u)}\geq e^{-2}$ and $\left[1-\left(1-(1-u^{\frac{1}{n_0}})p\right)^{n_0-1}\right]\geq(1-u^{\frac{1}{n_0}})\frac{n_0-1}{2K}\geq\frac{1}{4}(1-u^{\frac{1}{n_0}})$ for $K\geq 2$.

Remark: For $\rho_0=\frac{a}{L_0}$ for any a, $n_0=1+a$. Thus, $E_0(\rho_0,1,L_0)\geq\frac{(1-u^{\frac{1}{1+a}})a}{2K}$. In particular, with $a=1$, $E_0(\rho_0,1,L_0)\geq\frac{(1-u^{\frac{1}{2}})}{2K}$.

B. Order-Tight Results for Necessary and Sufficient Number of Tests with Group Testing

In this section, we present a brief sketch of the derivation of the order results for the necessary number of tests presented in Table I. We first note that $I^{(j)}=H(Y|X_{S(K-j)})-H(Y|X_{S(K-j)},X_{S(j)})$ [19], where $H(\cdot|\cdot)$ represents the entropy function [30]. From (2), we have

$$\begin{aligned}H(Y|X_{S(K-j)})=\sum_{l=0}^{K-j}\Bigg[\binom{K-j}{l}p^l(1-p)^{K-j-l}\\ \times H_b\left((1-q)u^l(1-p(1-u))^j\right)\Bigg]\end{aligned} \tag{44}$$

$$\begin{aligned}H(Y|X_{S(K-j)},X_{S(j)})=\sum_{i=0}^{K}\Bigg[\binom{K}{i}p^i(1-p)^{K-i}\\ \times H_b\left((1-q)u^i\right)\Bigg].\end{aligned} \tag{45}$$

We use the results from [35] for bounding the mutual information term. We collect the required results from [35] in the following lemma.

Lemma 2. *Bounds on $I^{(j)}$ [35]: Let $p=\frac{\delta}{K}$. $I^{(j)}$ can be expressed as $I_1^{(j)}+I_2^{(j)}$, where*

$$I_1^{(j)}=\delta e^{-\delta(1-u)}(1-q)\left(u\log u+1-u\right)\frac{j}{K}+O\left(\frac{1}{K^2}\right). \tag{46}$$

For the case with $u = 0$ and $q > 0$ we have:

$$I_2^{(j)} = \delta e^{-\delta}\left(\log(\frac{1}{q}) - (1-q)\right)\frac{j}{K} + O\left(\frac{1}{K^2}\right), \qquad (47)$$

and for $q = 0$, $u \geq 0$ we have:

$$\begin{aligned}\delta e^{-\delta}\left((1-u)\left[\log\frac{K}{j\delta(1-u)}\right] - u\right)\frac{j}{K} + O\left(\frac{1}{K^2}\right) \leq I_2^{(j)} \\ \leq \delta e^{-\delta(1-u^2)}\left((1-u)\left[\log\frac{K}{j\delta(1-u)}\right] - u + u^2\right)\frac{j}{K} \\ + O\left(\frac{1}{K^2}\right). \qquad (48)\end{aligned}$$

Thus, with $\delta = 1$ and large K, neglecting $O(1/K^2)$ terms, we get: (a) For $u = 0$, $q > 0$ case, $I^{(j)} \approx \frac{j}{eK}\log(\frac{1}{q})$. (b) For $q = 0$, $0 \leq u \leq 0.5$ case, simplifying further, we get

$$\frac{j}{eK}(1-u)\log\frac{K}{j} \lessapprox I^{(j)} \lessapprox \frac{j}{e^{1/2}K}(1-u)\left(\log\frac{K}{j} + 1\right). \qquad (49)$$

In the above, we have used the notation "$\approx$" and "$\lessapprox$" to highlight the fact that $O(\frac{1}{K^2})$ terms have been neglected in the above expressions for $I^{(j)}$. The order results for lower bounds now follow by first noting that $\max_{1\leq j\leq K}\frac{\Gamma_l(L,N,K,j)}{I^{(j)}} \geq \frac{\Gamma_l(L,N,K,1)}{I^{(1)}}$, and, for the scaling regimes under consideration the combinatorial term, $\Gamma_l(L,N,K,1)$ can be asymptotically bounded as $\lim_{N\to\infty}\Gamma_l(L,N,K,1) \geq \log\frac{1}{1-\alpha_0}$.

References

[1] A. Sharma and C. R. Murthy, "On finding a set of healthy individuals from a large population," in *Information Theory and Applications Workshop, San Diego, CA, USA*, 2013, pp. 1–5.

[2] R. Dorfman, "The Detection of Defective Members of Large Populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, Dec. 1943.

[3] D. Du and F. Hwang, *Pooling designs and non-adaptive group testing: Important tools for DNA sequencing*, World Scientific, 2006.

[4] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[6] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[7] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[8] D. Cabric, S. M. Mishra, D. Willkomm, R. Brodersen, and A. Wolisz, "A cognitive radio approach for usage of virtual unlicensed spectrum," in *Proc. of 14th IST Mobile Wireless Communications Summit*, 2005.

[9] FCC, "Et docket no. 02-155," *Spectrum policy task force report*, Nov. 2002.

[10] A. Sharma and C. R. Murthy, "Group testing-based spectrum hole search for cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3794–3805, Oct 2014.

[11] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 363–377, 1964.

[12] P. Erdos, P. Frankl, and Z. Furedi, "Families of finite sets in which no set is covered by the union of r others," *Israel Journal of Mathematics*, vol. 51, no. 1-2, pp. 79–89, 1985.

[13] A. G. Dyachkov and V. V. Rykov, "Bounds on the length of disjunctive codes," *Problems of Information Transmission*, vol. 18, no. 3, pp. 7–13, 1982.

[14] A. Sebo, "On two random search problems," *Journal of Statistical Planning and Inference*, vol. 11, no. 1, pp. 23–31, Jan. 1985.

[15] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli, "Group testing with probabilistic tests: Theory, design and application," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 7057–7067, Oct. 2011.

[16] M. B. Malyutov, "The separating property of random matrices," *Mat. Zametki*, vol. 23, no. 1, pp. 155–167, 1978.

[17] M. B. Malyutov, "On the maximal rate of screening designs," *Theory Probab. and Appl.*, vol. 24, pp. 655–667, 1979.

[18] M. B. Malyutov and P. S. Mateev, "Planning of screening experiments for a nonsymmetric response function," *Mat. Zametki*, vol. 27, no. 1, pp. 109–127, 1980.

[19] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, 2012.

[20] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3019–3035, May 2014.

[21] J. Scarlett and V. Cevher, "Phase transitions in group testing," in *Proc. Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, Jan. 2016, pp. 40–53.

[22] M. Ruszinkó, "On the upper bound of the size of the r-cover-free families," *J. Comb. Theory, Ser. A*, vol. 66, no. 2, pp. 302–310, 1994.

[23] A. Emad and O. Milenkovic, "Poisson group testing: A probabilistic model for boolean compressed sensing," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4396–4410, Aug 2015.

[24] A. G. D'yachkov, "Bounds for error probability for a symmetrical model in designing screening experiments," *Problems Inform. Transmission*, vol. 17, pp. 245–263, 1981.

[25] Hung Q. Ngo, Ely Porat, and Atri Rudra, "Efficiently decodable error-correcting list disjunct matrices and applications - (extended abstract).," in *ICALP (1)*, Luca Aceto, Monika Henzinger, and Jir Sgall, Eds. 2011, vol. 6755 of *Lecture Notes in Computer Science*, pp. 557–568, Springer.

[26] J. Scarlett and V. Cevher, "How little does non-exact recovery help in group testing?," in *Proc. ICASSP*, March 2017, pp. 6090–6094.

[27] A. Sharma and C. R. Murthy, "Computationally tractable algorithms for finding a subset of non-defective items from a large population," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7149 – 7165, Nov. 2017.

[28] J. Yoo, Y. Xie, A. Harms, W. U. Bajwa, and R. A. Calderbank, "Finding zeros: Greedy detection of holes," *eprint arXiv:1303.2048*, 2013.

[29] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York, NY, USA, 1968.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, NY, USA, 1991.

[31] G. Atia and V. Saligrama, "A mutual information characterization for sparse signal processing," in *The 38th International Colloquium on Automata, Languages and Programming (ICALP), Switzerlnd*, 2011.

[32] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and nonlinear observations: A unified shannon-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 749–776, Feb 2017.

[33] A. C. Gilbert and M. J. Strauss, "Analysis of data streams: Computational and algorithmic challenges," *Technometrics*, vol. 49, no. 3, pp. 346–356, 2007.

[34] A. J. Macula and L. J. Popyack, "A group testing method for finding patterns in data," *Discrete Appl. Math.*, vol. 144, no. 1-2, pp. 149–157, 2004.

[35] D. Sejdinovic and O. Johnson, "Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction," in *Proc. Allerton Conf. on Commun., Control and Comput.*, 2010, pp. 998–1003.